**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное автономное образовательное учреждение**

**высшего образования**

**«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

КАФЕДРА АНАЛИЗА ДАННЫХ И

ТЕХНОЛОГИЙ ПРОГРАММИРОВАНИЯ

Направление: 09.03.03 – Прикладная информатика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**Разработка приложения для проектирования ИИ моделей для участия в гоночной игре Trackmania с использованием машинного обучения**

**Работа завершена:**

Студент 4 курса

группы 09-052

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Близнюк В.Ю.

**Работа допущена к защите:**

Научный руководитель

ученая степень, ученое звание,

должность

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сабитов Ш.Р.

Заведующий кафедрой

к.ф.-м.н.

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Бандеров В.В.

Казань-2024

Содержание

[Аннотация 3](#_Toc167050490)

[Введение 5](#_Toc167050491)

[1. Исследование предметной области создания искусственного интеллекта для гоночных игр 7](#_Toc167050492)

[1.1. Анализ предметной области 7](#_Toc167050493)

[1.2. Обзор искусственного интеллекта в большинстве гоночных видеоигр на примере серии видеоигр Need For Speed 8](#_Toc167050494)

[1.3. Обзор драйватаров в серии видеоигр Forza 9](#_Toc167050495)

[1.4. Обзор механики призраков в видеоигре TrackMania 11](#_Toc167050496)

[1.5. Обзор TMRL 12](#_Toc167050497)

[1.6. Описание программных средств реализации 12](#_Toc167050498)

[2. Проектирование приложения для создания моделей 14](#_Toc167050499)

[2.1. Составление технического задания 14](#_Toc167050500)

[3. Разработка приложения для обучения и исполнения моделей 15](#_Toc167050501)

[3.1. Реализация связи приложения и видеоигры 15](#_Toc167050502)

[3.2. Создание gym окружения видеоигры для обучения 19](#_Toc167050503)

[3.3. Реализация нейронной сети 21](#_Toc167050504)

# Аннотация

Abstract

# Введение

Область разработки искусственного интеллекта и машинного обучения активно развивается. С каждым днём появляются новости о новых языковых моделях, улучшенных алгоритмах обучения роботов или дронов. В условиях стремительного развития технологий и увеличения интереса к искусственному интеллекту, создание систем, способных эффективно взаимодействовать с игровыми средами, становится важной задачей. На данный момент внутриигровые соперники ограничиваются либо ботами со строго прописанными правилами, либо рекордами игроков. Оба подхода имеют свои минусы. Неживые соперники не могут адаптироваться к новым условиям и не предоставляют ощущения соревновательности, которое предлагает гонка с живыми игроками. Записанные рекорды дают это чувство, но сражение происходит не с человеком, а с его «призраком», который только повторяет действия игрока. И это не говоря про возможность столкновения с машиной соперника, что не представляется возможным из-за нематериальности «призрака».

На сегодняшний день существует небольшое количество исследований и приложений, посвященных разработке искусственного интеллекта для гоночных игр, которые могли бы решить приведенные проблемы. Например, Playground Games с их серией видеоигр Forza, используют механику драйватаров, в которой анализируется поведение игроков, после чего неживые противники пытаются имитировать то, как водит соответствующий боту человек.

Однако искусственный интеллект, основанный на нейронных сетях, чаще всего встречается в пользовательских проектах. Однако и они обладают своими недостатками, основной из которых – отсутствие доступа к состоянию игры, из-за чего входные данные нейронной сети приходится получать из частых снимков экрана запущенной игры, а сама видеоигра получает входные данные с помощью имитации использования клавиатуры или геймпада.

Актуальной данной работы заключается в том, что на сегодняшний день не существует приложений, которые могли бы имитировать поведение реального игрока в гоночные видеоигры, то есть искусственный интеллект которых был бы основан на использовании нейронных сетей, при этом имея доступ к состоянию видеоигры и возможность отдавать приказы напрямую, а не через эмуляцию устройства ввода. В процессе консультации с научным руководителем была поставлена цель и выделены задачи. Была налажена регулярная коммуникация с научным руководителем для уточнения вопросов и консультаций.

Целью ВКР является создание приложение для создания моделей искусственного интеллекта с помощью машинного обучения, которые могут быть использованы в гоночной игре TrackMania Nations Forever.

Задачи выпускной квалификационной работы:

1. Провести анализ предметной области;
2. Провести обзор существующих аналогов;
3. Составить техническое задание к функциональной части;
4. Описать программные средства реализации приложения;
5. Приготовить видеоигру и средства ввода и получения данных;
6. Разработать нейронную сеть и скрипт обучения;
7. Протестировать полученные модели против реальных игроков.

Выпускная квалификационная работа состоит из 4-ех глав, введения, заключения, списка использованной источников и приложений. В работе описан весь процесс проектирования, конструирования и разработки приложения, представлены скриншоты результатов обучения моделей нейронной сети, описания классов, а также вставлены наиболее важные части кода приложения.

# 1. Исследование предметной области создания искусственного интеллекта для гоночных игр

# 1.1. Анализ предметной области

При анализе существующих решений были выделены закономерности подходов к созданию искусственного интеллекта в гоночных видеоиграх. Большинство игр имеют интеллект соперников, напрямую прописанный в коде, никак не обучающийся и не адаптирующийся, то есть не использующий нейронные сети. Использование подобных средств редко и часто не реализует весь потенциал машинного обучения. Помимо этого, четко фиксированные правила сложны для разработки, так как, в зависимости от сложности физической модели видеоигры, нужно учитывать множество различных параметров: от положения и поворота машины на трассе до состояния дорожного покрытия и погоды.

Если же разработчики видеоигры используют машинное обучение, то часто оно ограничено или до подсказок самому игроку в том, как лучше проехать тот или иной поворот, или до незаметной помощи ботам. Во втором случае, однако, это часто имеет вспомогательный и подсказывающий характер, который с некоторым шансом влияет на решение, принятое игрой на основе предписанных правил.

Пользовательские и фанатские проекты же, которые используют нейронные сети, имеют проблему в виде невозможности напрямую получать состояние видеоигры или отдавать ей команды для движения автомобиля. Чтобы иметь входные данные для обучения и использования модели, приходится делать десятки снимков экрана в секунду и пытаться их обработать, чтобы модель получала только важные и необходимые данные. Очевидно, это одновременно ресурсозатратно и сложно для реализации. Данные, полученные со скриншота, могут быть неполными или обработанными неправильно, что также могут повлиять на работу искусственного интеллекта. Второй проблемой является ввод действий в игру, что часто приходится имитировать с помощью эмуляции устройства ввода и его нажатий.

Некоторые видеоигры, которые предполагают соревнование в том, кто проедет трассу быстрее, а не первее, такие как TrackMania, часто имеют механику призраков – предзаписанных другим игроком действий того, как он сам проходил трек. Их проблема заключается в том, что такие соперники должны создаваться отдельно для каждой карты, машины или другого фактора. Изменение даже одного из них может повлиять на поведение призрака и сделать его несостоятельным. Кроме того, это не позволяет выбрать сложность соперника, если трасса новая или непопулярная, и не дает возможности сталкиваться с машиной другого гонщика.

# 1.2. Обзор искусственного интеллекта в большинстве гоночных видеоигр на примере серии видеоигр Need For Speed

Искусственный интеллект большинства видеоигр создан с помощью большого и четкого набора правил, которых придерживаются внутриигровые соперники. Так было с создания самых первых игр этого жанра, так как на тот момент использование нейронных сетей было мало распространено и не было эффективным с производительность персональных компьютеров и консолей тех лет.

Из-за этого для каждой ситуации на дороге в коде видеоигр прописано то, как должен действовать гонщик. Положение каждой машины на трассе, их скорости и модели, текущая погода, всё это влияет на поведение машины и прописано сложной системой условных выражений. Конечно, создавать такую систему очень сложно, так как нужно учитывать все вышеперечисленные факторы, влияющие на то, как будет реагировать на ввод машина.

Так, на рисунке 1.1 машины на 192 и 193 местах выбирают то, как нужно действовать, не на основе своих знаний или благодаря данным от обучения, а от предписанных в коде правил.



Рисунок 1.1. Скриншот из видеоигры Need For Speed: The Run

# 1.3. Обзор драйватаров в серии видеоигр Forza

PlayGround Games с их серией видеоигр Forza решили шагнуть дальше и уже в первой своей видеоигре Forza Motorsport, выпущенной в 2005 году, ввели механику драйватаров. Во время прохождения трассы игра изучала поведение игрока и использовала полученные данные, чтобы создать внутриигрового соперника со стилем вождения, как у реального человека.

Тогда драйватары сохранялись локально, и чтобы сразиться со своим другом, приходилось копировать файлы сохранения. Это изменилось с выходом Forza Motorsport 5 в 2013 году, когда данные о драйватарах сохранялись в облаке. Это позволило участвовать в гонках сразу с несколькими соперниками, которые хоть и управлялись компьютером, но пытались имитировать поведение соответствующих себе живых игроков.

Однако система не является совершенной. Драйватары не используются для непосредственного принятия решения на трассе, а лишь дают подсказки обыкновенному искусственному интеллекту. Несовершенство такого решения также подтверждается тем, что даже в последних играх серии сложность влияет не только на то, какие драйватары будут использованы в гонке, но и на их автомобили, искусственно замедляя или ускоряя соперников в зависимости от сложности и игровой ситуации.

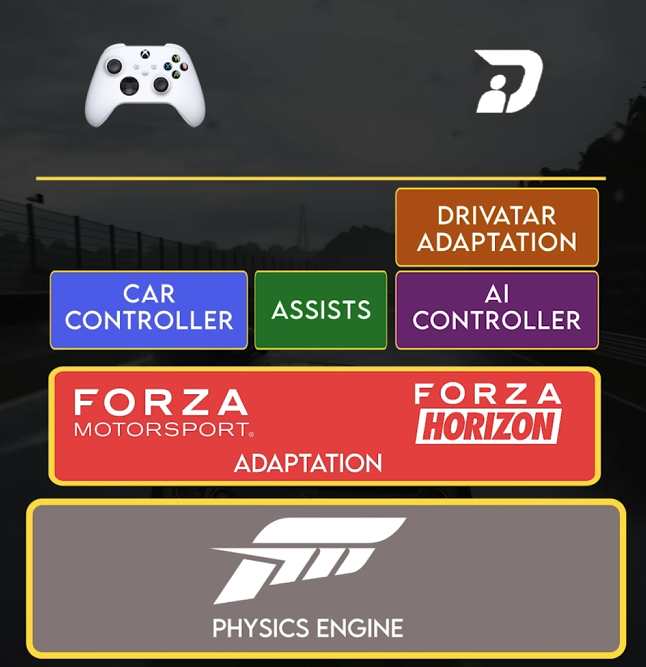


Рисунок 1.2. Система управления автомобилем в играх серии Forza

# 1.4. Обзор механики призраков в видеоигре TrackMania

Nadeo в своей видеоигре TrackMania совсем отказались от физических соперников на трассе, создав игру для того, чтобы игроки соревновались в рекордах по времени прохождения, а не в том, кто кого обогнал.

Такие рекорды позволяют получить чувство соревнования с другими живыми игроками, однако они не появляются на трассе в виде физической машины с коллизиями, что портит ощущение погружения в игровой процесс. Другой проблемой является то, что рекорды должны быть записаны конкретными людьми на конкретных трассах в конкретных условиях. Очевидно, что в такой ситуации выбор сложности для непопулярных и новых карт не представляется возможным.



Рисунок 1.3. Демонстрация призраков в гонке TrackMania Nations Forever

# 1.5. Обзор TMRL

TMRL – пользовательский фреймворк для создания приложений для обучения с подкреплением, которое идет со встроенными возможностями подключения фреймворка к видеоигре TrackMania 2020.

На странице проекта демонстрируются возможности фреймворка на примере создания рекордов, способных проходить трассы различной сложности и правильно оценивающих окружающую обстановку.

Однако, данные фреймворк получает с помощью последовательных снимков экрана и их последующей обработки, а ввод в игру имитируется эмуляцией геймпада, из-за чего программу во время обучения или выполнения нельзя сворачивать, а скриншоты и их обработка могут давать ложные результаты.

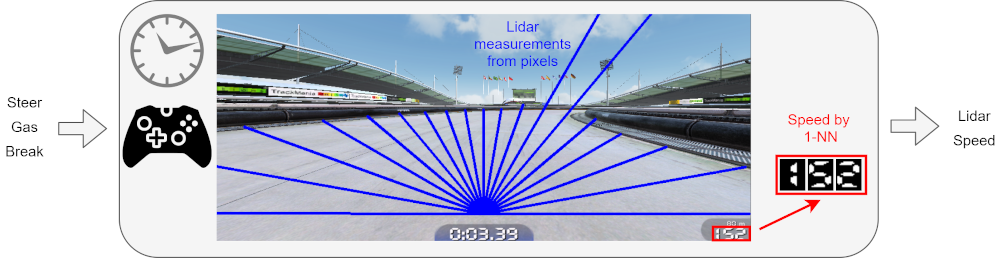


Рисунок 1.4. Демонстрация получения и ввода данных в TMRL

# 1.6. Описание программных средств реализации

В качестве видеоигры, для которой приложение будет создавать модели, была выбрана TrackMania Nations Forever, так как игра имеет детерминистскую машину состояний и четкую физическую модель, что позволяет модели нейронной сети обучаться без опасения натолкнуться на какое-то случайное событие. Для получения данных игры и ввода в неё было решено использовать TMInterface. Для работы с ней будет использована соответствующая Python библиотека.

В качестве языка разработки был выбран Python, а библиотеки для создания нейронной сети – Torch. Для создания окружения для нейронной сети использована библиотека gym.

# 2. Проектирование приложения для создания моделей

# 2.1. Составление технического задания

В результате анализа предметной области, цели и поставленных задач возникла необходимость разработки приложения, позволяющего создавать и сохранять модели нейронной сети, способной проходить трассы в гоночной видеоигре TrackMania Nations Forever. Всё приложение должно быть написано на языке программирования Python. Связь приложения с видеоигрой должна осуществляться с помощью TMInterface, окружение для нейронной сети должно быть создано с помощью библиотеки gym, нейронная сеть должна быть написана с помощью библиотеки Torch.

Приложение должно представлять собой программу, которая при введенных данных о треке и, при необходимости, сохраненных данных обучения, должна начать процесс обучения нейронной сети для участия в гонке и её прохождения. При наличии сохраненных данных обучения программа также должна иметь возможность загрузить модель и использовать ее для прохождения трассы.

Обучение должно выполняться автоматически и без необходимости участия человека или открытого окна видеоигры или программы. Созданные модели должны уметь проходить трассы, прохождение которых человеком занимает 15 секунд, за время, не превышающее 30 секунд. Обучение модели, достигающей такого результата, должно занимать не более 8 часов. Приложение должно давать возможность получить данные от треке при указании пути до файла, которых должно быть достаточно для того, чтобы ввести их при начале обучения или исполнения модели.

# 3. Разработка приложения для обучения и исполнения моделей

# 3.1. Реализация связи приложения и видеоигры

Исходя из технического задания, были написаны необходимые классы и интерфейсы, которые обеспечивали доступ из приложения к состоянию видеоигры и возможность ввода для управления машиной.

Использовалась пользовательская модификация видеоигры TrackMania Nations Forever под названием TMInterface, которая позволяет наладить связь с собой с помощью одноименной библиотеки для Python.

Основными классами библиотеки, которые использовались в приложении, являются:

* + Client – класс, напрямую связанный с клиентом игры. Предназначен в первую очередь для наследования от себя и переопределения методов обратного вызова.
  + TMInterface – основной класс, который позволяет получать состояние игры или менять его, а также взаимодействовать с самими процессами видеоигры.
  + SimStateData – класс, который хранит в себе состояние видеоигры.
  + HmsDynaStruct – класс, который хранит в себе состояние об автомобиле в конкретный момент времени.
  + CheckpointData – класс, который хранит в себе состояние о пройденных чекпоинтах.

Для наследования от класса Client был создан класс SimStateClient. Он хранит в себе поля, отвечающие за хранение ссылки на объекты классов TMInterface, SimStateData и CheckpointData, данные, которые должны быть введены в игру, факт завершения гонки и флаг, при поднятии которого клиент должен перезапустить гонку.

Было переопределено 2 метода класса Client:

* + on\_run\_step – вызывается каждый физический тик в гонке, имеет в параметрах TMInterface, который хранит все текущие состояния. Метод используется для сохранения в клиенте ссылок на TMInterface, SimStateData и CheckpointData. Также при наличии ввода передает его в игру, а при необходимости перезапуска гонки – перезапускает её.
  + on\_checkpoint\_count\_changed – вызывается при прохождении чекпоинта в гонке, в том числе финишной черты. Используется для того, чтобы гонка не завершалась при проезде финиша, при этом меняя поле класса, отвечающее за факт конца гонки.

Также был создан класс SimStateInterface, обеспечивающий связь с клиентом. При инициализации он создает по экземпляру классов SimStateClient и TMInterface, регистрирует интерфейс в клиенте, а также хранит обработанные данные о состоянии машины. С помощью своих методов он также позволяет обращаться к клиенту при необходимости ввода каких-либо данных или для их получения.

Для хранения обработанных данных о машине был создан класс SimState. Он имеет в себе поля, которые уже передаются в нейронную сеть. Основные хранимые данные:

* + linear\_speed – линейная скорость автомобиля, то есть скорость по основной оси машины.
  + angular\_speed – боковая скорость автомобиля, то есть скорость по оси машины, перпендикулярной основной.
  + distance\_to\_centerline – расстояние до центра дороги.
  + angle\_to\_centerline – угол основной оси машины относительно дороги, где 0 означает, что машина смотрит в сторону, в которую нужно двигаться, а 1 по модулю – что машина смотрим в обратную сторону.
  + next\_curve\_distance – расстояние до ближайшего поворота.
  + next\_curve\_direction – направление следующего поворота (-1 – влево, 1 – вправо)
  + next\_curve\_distance2, next\_curve\_distance3 – длина следующих прямых отрезков дороги.
  + next\_curve\_direction2, next\_curve\_direction3 – направление следующих поворотов.

Помимо сложных методов подсчета этих данных класс также нормализует их до отрезка от -1 до 1.

Во время разработки было создано несколько сущностей подобных классов, отвечающих за разные типы трасс. Самый простой для трека, состоящего из одного прямого участка прямо до финиша, принимал только первые 4 указанных поля, так как не было необходимости знать, сколько осталось до поворотов и их направление по причине их отсутствия.

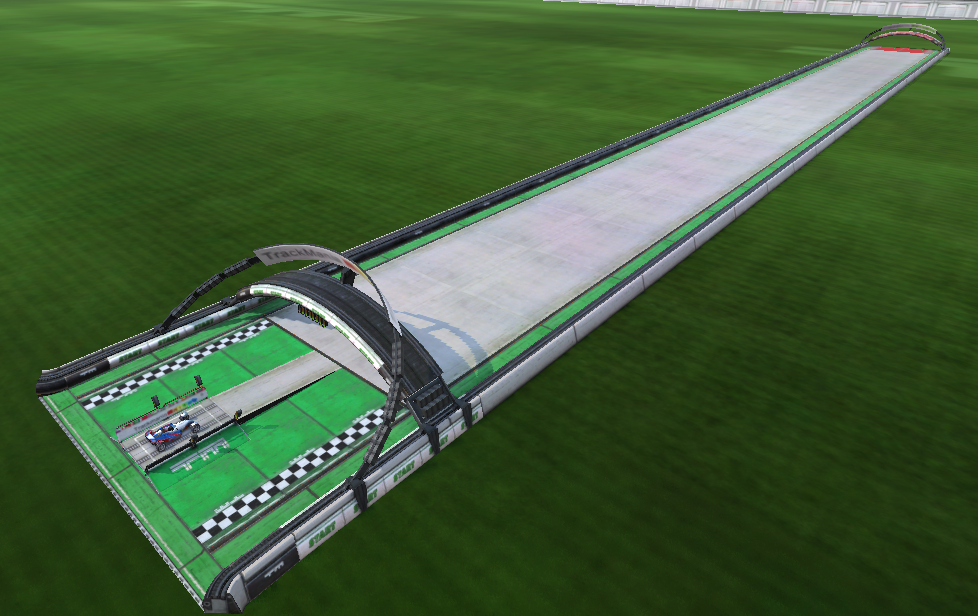


Рисунок 3.1. Прямая трасса

Класс, позволяющий получать данные для карты вида «змейка», имел в себе первые 6 указанных параметров, то есть теперь имел в себе значения, позволяющие понять, когда нужно поворачивать и в какую сторону.



Рисунок 3.2. Карта вида «змейка»

Самый большой класс для карт с различными перпендикулярными поворотами, имеет все 10 указанных полей, чтобы нейросеть понимала, стоит ли сразу разгоняться после поворота или нужно притормозить, так как скоро придется поворачивать ещё.



Рисунок 3.3. Карта с перпендикулярными поворотами

Для удобства разработки, средствами TMInterface были включены отображение текущего ввода и основные данные об автомобиле



Рисунок 3.4. Интерфейс запущенного TMInterface

# 3.2. Создание gym окружения видеоигры для обучения

Для обучения нейронной сети было решено создать окружение с помощью библиотеки gym. gym – это библиотека для разработки и тестирования алгоритмов обучения с подкреплением. Она предоставляет различные окружения и возможности создания окружений, в которых агенты могут обучаться и взаимодействовать, что позволяет исследователям и разработчикам создавать, тестировать и сравнивать различные алгоритмы обучения с подкреплением.

Библиотека предоставляет класс Env, предполагаемый для наследования и переопределения методов обратного вызова. В связи с этим был создан класс TMEnv. Основные методы окружений gym:

* + step – метод, который вызывается каждый шаг алгоритма обучения и обычно вызывает остальные методы обратного вызова.
  + render – метод, обычно используемый для отображения каких-либо данных во время обучения.
  + reset – метод, предназначенный для приведения состояния окружения в свое обычное состояние.
  + reward – метод, возвращающий подкрепление, или награду, которое используется в алгоритме обучения.

Этот класс инициализирует и содержит в себе экземпляр класса SimStateInterface, что позволяет получать актуальные данные из состояния видеоигры и вводить действия для машины.

Самой сложной для создания являлась функция награды, так как от нее зависит то, как именно нейронная сеть будет менять веса. Было решено давать подкрепление за высокую скорость, высокое ускорение, прохождение трассы в правильном направлении, прохождение поворотов и трека в целом. Отрицательная награда даётся за отрицательную скорость, за непрохождение трассы или прохождение трассы и поворотов в неправильном направлении, за падение из трека и за факт слишком сильного поворота машины в неправильном направлении.

Другой важной функцией является функция step, вызываемая каждый шаг обучения. Она принимает ввод для видеоигры, получаемый от модели, обрабатывает его, вводит в игру, сохраняет полученные данные, проверяет, не нужно ли перезапустить обучение и перезапускает при необходимости, а также возвращает все необходимые для нейронной сети данные:

* + observation – обработанные и нормализованные данные, полученные из клиента (SimStateClient), интерфейса (SimStateInterface) и класса обработки состояния (SimState) видеоигры.
  + reward – награда, полученная алгоритмом обучения за текущий шаг.
  + done – булевая переменная, означающая, закончилась ли текущая сессия обучения и было ли перезапущено окружение.

Помимо этого, TMEnv также хранит множество вспомогательных параметров, большинство из которых позволяют функции награды более корректно определять подкрепление, как, например, previous\_distance\_to\_curve, отвечающая за расстояние до поворота в предыдущий шаг обучения и с которой сравнивается текущее расстояние до поворота.

Этот класс также обеспечивает факт запущенности видеоигры, устанавливает ее скорость, хранит в себе значения, используемые нейронной сетью. Основная из таких - observation\_space, определяющая, сколько входных и параметров будет принимать модель нейронной сети. Она имеет определение в виде

OBSERVATION\_SPACE = Box(-1, 1, (10,)).

Это означает, что нейросеть принимает 10 параметров, все из которых принимают значения в диапазоне от -1 до 1.

# 3.3. Реализация нейронной сети

Для обучения с подкреплением было решено использовать метод обучения DQN (Deep Q-Network). Его основными компонентами являются:

* + Q-значение – Q(s,a) предствавляет собой ожидаемую суммарную награду, которую агент получит, начиная из состояния s, выполняя действие a и следуя некоторой стратегии в будущем.
  + Нейронная сеть – для аппроксимации функций Q-значений используется глубокая нейронная сеть. Входом сети является состояние s, а на выходе сеть выдает Q-значения для всех возможных действий a в этом состоянии.
  + Целевая сеть – DQN использует две сети: основную (policy) сеть и целевую (target) сеть. Целевая сеть копирует веса основной сети через определенные интервалы времени. Это помогает стабилизировать обучение, так как целевые значения изменяются медленнее.
  + Память воспроизведения опыта (Replay Memory) – опыт агента (переходы) сохраняется в буфере воспроизведения. Переходы случайным образом выбираются из буфера для обучения. Это снижает корреляцию между последовательными шагами и делает обучение более стабильным и эффективным.

Далее каждый из компонентов будет описан более подробно.

В памяти воспроизведения опыта (Replay Memory) хранятся переходы, которые наблюдает агент, что позволяет повторно использовать эти данные в дальнейшем. Путем случайной выборки из нее переходы, составляющие партию, декоррелируются, что значительно стабилизирует и улучшает процедуру обучения сети.

Для этого понадобилось 2 класса:

* + Transition – именованный кортеж, представляющий один переход в окружении. Он сопоставляет пары (состояние, действие) с их результатом (следующее\_состояние, награда).
  + ReplayMemory – циклический буфер ограниченного размера, в котором хранятся переходы, наблюдавшиеся в последнее время. Он также реализует метод sample для выбора случайной партии переходов для обучения.

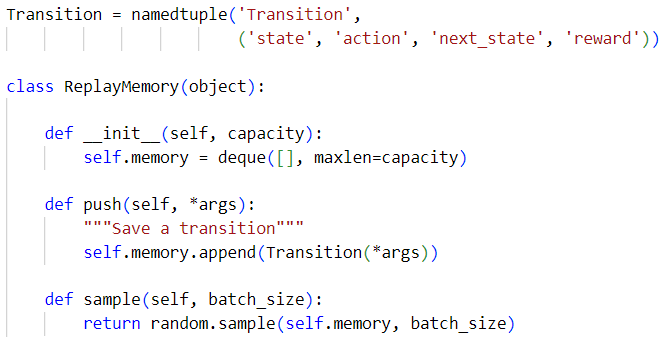


Рисунок 3.5. Реализация классов памяти воспроизведения опыта

Основной целью алгоритма обучения с подкреплением является обучение политики, которая пытается максимизировать дисконтированное кумулятивное вознаграждение

,

где дисконт γ является константой от 0 до 1. Низкие значения дисконта делают вознаграждения из неопределенного далекого будущего менее важными для агента, чем те, которые он может получить в ближайшем будущем и в которых он может быть достаточно уверен. Они также побуждают агентов получать вознаграждения ближе по времени, чем эквивалентные вознаграждения, которые находятся в будущем.

Основная идея Q-обучения заключается в том, что если бы была функция Q\*: State×Action→R, которая могла бы сказать, какой будет награда, если будет предпринято действие в данном состоянии, то можно было бы легко построить политику, которая максимизирует вознаграждение:

.

Однако агент не знает всего о мире, поэтому у него нет доступа к Q\*. Но поскольку нейронные сети являются универсальными аппроксиматорами функций, можно просто создать такую сеть и обучить ее, чтобы она напоминала Q\*.

Для правила обновления обучения было решено использовать факт, что каждая функция Q для некоторой политики подчиняется уравнению Беллмана:

.

Разность между двумя сторонами равенства называется погрешностью временной разности, δ:

.

Чтобы минимизировать эту ошибку, мы будем использовать потерю Хьюбера. Функция потерь Хьюбера действует как средняя квадратичная ошибка, когда ошибка мала, и как средняя абсолютная ошибка, когда ошибка велика – это делает его более устойчивым к выбросам, когда оценки Q очень зашумлены. Это вычисляется для пакетов переходов B, взятых из памяти воспроизведения опыта:

,

где

Модель представляет собой нейронную сеть с прямой передачей данных, которая учитывает разницу между текущим и предыдущим состояниями. У нее есть два выхода, представляющие собой Q(s,left) и Q(s,right) (где s - вход сети). По сути, сеть пытается предсказать ожидаемую прибыль от каждого действия с учетом текущих входных данных.

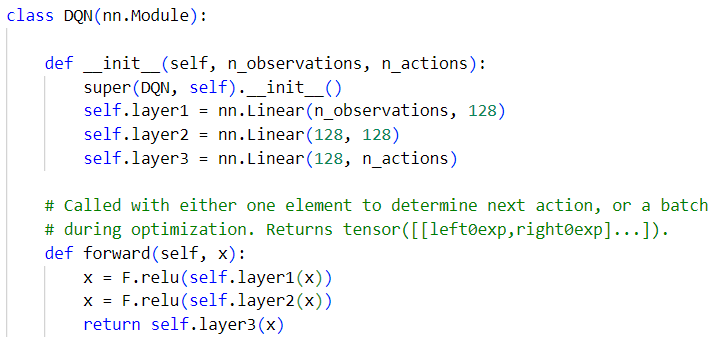


Рисунок 3.6. Реализация класса нейронной сети

Скрипт обучения имеет много различных параметров, основные из которых было решено выделить в отдельный класс, как гиперпараметры.

* + BATCH\_SIZE = 128 – размер пакета переходов, которые выбираются из памяти воспроизведения опыта для каждого шага обучения. Большой размер пакета может стабилизировать обучение за счет более точной оценки градиента, но также требует больше памяти и вычислительных ресурсов. Выбранный размер пакета в 128 является распространенным выбором, который балансирует эти факторы.
  + GAMMA = 0.99 – дисконтирующий фактор для подкрепления, описанный ранее.
  + EPS\_START = 0.95 – начальное значение эпсилон в эпсилон-жадной политике. Эпсилон управляет балансом между исследованием (пробованием новых действия) и эксплуатацией (выбором наилучшего известного действия). Высокое начальное значение эпсилон означает, что агент будет много исследовать в начале, пробуя разные действия для сбора информации об окружении.
  + EPS\_END = 0.05 – конечное значение эпсилон после его снижения. Это значение, до которого эпсилон уменьшится со временем. Низкое конечное значение эпсилон означает, что в конечном итоге агент будет в основном использовать то, что он уже выучил, но иногда будет исследовать, чтобы избежать застревания в субоптимальном поведении.
  + EPS\_DECAY = 2500 – скорость, с которой эпсилон уменьшается от EPS\_START до EPS\_END.
  + TAU = 0.005 - скорость обновления целевой сети. Параметр τ управляет тем, как быстро целевая сеть обновляется, чтобы соответствовать основной сети. Малое значение τ означает, что целевая сеть обновляется медленно, что приводит к более стабильному обучению.
  + LR (Learning Rate) = 0.0001 - скорость обучения оптимизатора, конкретно AdamW. Скорость обучения определяет размер шагов, которые делает оптимизатор в направлении градиента. Меньшая скорость обучения означает, что агент будет делать более точные обновления параметров сети, что может привести к более стабильному и точному обучению, но это может также замедлить сходимость.

Основной функцией выбора действия является функция select\_action, которая выбирает действие в соответствии с эпсилон-жадной политикой.

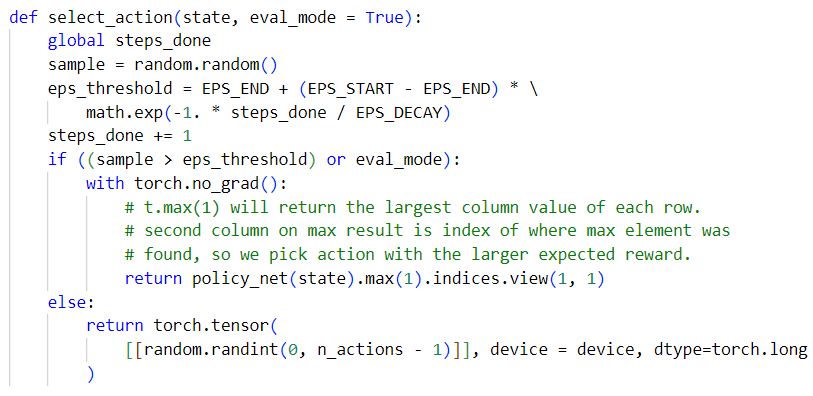


Рисунок 3.7. Реализация функции select\_action